DOCUMENT RESUME

ED 126 141 TM 005 397

AUTHOR Hall, Mary
TITLE Dissemination and Utilization of Large Scale Test Results: Is Anybody Out There Listening? And If Not, Why Not?
PUB DATE [Apr 76]
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Decision Making; Educational Assessment; Educational Policy; Elementary Secondary Education; *Information Dissemination; *Information Needs; *Information Utilization; National Programs; *Program Effectiveness; School Districts; Schools; State Programs; Testing Problems; *Testing Programs; *Test Results

ABSTRACT

The paper reviews previous research studies and conferences which have dealt with the question of whether large-scale testing programs are effective. It is concluded that such programs, defined as efforts to determine the status of student achievement on a school, district, state, or national basis, are not serving the informational needs of the decision-making bodies for whom they are designed. Three schools of thought are discussed concerning reasons why large-scale testing programs are not adequately responsive. These included those who believe that policymakers do not wish to make data-based decisions; those who believe the fault lies with ineffective dissemination and utilization subsystems; and those who challenge the suitability of large-scale testing programs, as currently operated, for serving the realities of educational policymaking. After discussing the nature of educational policymaking, the paper suggests three reasons why testing and assessment programs have failed to make the desired impact. These include: (1) such programs have not adequately defined the level at which their target audiences are most likely to make policy; (2) such programs seldom have the capacity to produce information which is "issue" oriented at a time when it is most needed by policymakers; and (3) few programs take into account that the policymaking process is characterized by "uncertainty" and by "competing value systems." (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED126141

DISSEMINATION AND UTILIZATION OF LARGE-SCALE TESTS RESULTS: IS ANYBODY OUT THERE LISTENING? AND IF NOT, WHY NOT?

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A paper to the American Educational Research Association

San Francisco, California
April 1976

TM005 397

By

Mary Hall
Associate Superintendent
Oregon State Department of Education

At past meetings of the American Educational Research Association, a variety of papers and symposia have dealt with technical issues related to the improvement of large-scale testing programs--those programs which we are defining as efforts to determine the status of student achievement on a school, district, state or national basis. We know that such programs are increasing in number and cost (Hawke, 1975). Their existence is commonly found as a prerequisite in the rhetoric for accountability in public education.

It appeared to many of us, however, that presentations during the last few years have increasingly ended by asking two basic questions:

- Are such programs actually producing data which serves the informational needs of the decision-making bodies for whom they are designed? Or to put it another way, is there any evidence that decisions, particularly of the policy nature, are being stimulated or enhanced by the data being produced by such programs?
- And if such programs are not having the desired impact, why not? What can be done to improve this situation?

This symposium was organized specifically to focus on these two questions.

In answering the first question, it is my belief that there is now sufficient empirical evidence and a sufficient number of public declarations of subjective evaluations by both producers and consumers of large-scale testing data to admit that such programs are not meeting the decision-making needs of their target audiences. Let me quickly review some of this evidence.

- Last year I presented the results of a nationwide survey of state assessment programs (Hall, 1975). One of the major findings of this study was that, with the exception of setting priorities for the allocation of funds under Title III of the Elementary and Secondary Education Act, less than a third of the states could "show evidence of using their assessment results to make the types of decisions which are frequently cited as a justification for the initiation of statewide assessment programs" (p. ii).

- In the evaluation of the assessment program of a state that has provided national leadership in the so-called "accountability movement," the authors declared that one of their most surprising findings was that the statewide assessment program had "little apparent value to any major group" (House, Rivers & Stufflebeam, 1974, p. 664).

- Similar complaints about the lack of utility of statewide assessment programs, at least in the policy arena, were voiced at the 1975 National Forum for the Advancement of State Educational Assessment Programs and the 1974 conference of the National Association of State Boards of Education. The harshest comments were made by state legislators (or their staffs), by chief state school officers or by members of state educational boards. Remember, these are the groups frequently cited as target audiences for state assessment data.

- In 1973, the Oregon Department of Education conducted an inventory and analysis of standardized norm-reference tests administered in the state on a district- or school-wide basis (deJung, 1973). One portion of the study dealt with the use made by the schools of their test data. Nearly all reported that the target audiences for such test results were teachers and that such tests were used either for classroom planning or for assigning student grades. Fewer than five percent reported that the data was used for any policy purposes at the building or district level. Yet, at the Department of Education, we are daily receiving communications from teachers about a new Minimum School Standard which requires that districts initiate program assessment activities. These teachers report that the current tests being used by their districts or buildings are "useless," so why mandate further activity in this area.

- Bleecher (1975) recently reported the results of a study in Michigan focusing on teacher attitudes towards accountability and assessment efforts at either the state or local level. His teacher respondents reported, among other findings, that assessment information produced from whatever source (state or local) and transmitted to teachers resulted in (a) recommendations with which they felt they could not physically or mentally comply,

> and (b) information which was irrelevant to their personal interests. Given these two findings, one might assume that these assessment programs had little impact on policymaking at the classroom level--at least to the extent such policies were chosen or enacted by teachers.

These are some of the indicators which have led me personally to the conclusion that large-scale testing programs, as presently conducted, have minimal or no impact on the educational policymaking audiences for whom they are supposedly designed. They are sustained only by a life-support system which is connected to the tenuous cord of federal and state requirements or a host of good intentions. I also believe that we must either improve this situation or that one of these days budget-conscious policymakers are going to use the data to make a decision that I am certain none of us are after. They are going to decide that the body of large-scale testing is no longer breathing and pull the plug.

There are at least three major schools of thought as to why large-scale testing programs are not adequately responsive to educational policymakers at the national, state or local level.

The first school of thought says the problem rests with the policymakers themselves. They either do not want to make any decisions or they do not want to make any data-based decisions. Those of you who advocate this position will get a certain amount of argument from policy scholars in other fields. Lindbloom (1968) for example quotes a former director of the federal budgeting system who admonished:

> The cynical view of the matter is that rational calculation in government programming is a harmless but ineffectual pursuit, since all

> important questions are ultimately decided on "political" grounds. . . .The thesis is wrong if it is taken to mean the findings of skilled and objective analysis of public programs are not influential in decision-making at the highest level. In fact, such findings are usually influential and, not infrequently, decisive (p. 11).

There is also evidence from a recent study of the Institute for Social Research at the University of Michigan (1975) that top-level federal policymakers are looking for help from social scientists and are using the results of social science research to shape policy decisions. As quoted in Education Daily:

> . . .Our data suggest that government executives do not need to be sold on the potential usefulness of scientific information, nor do they lack a reasoned appreciation of its value in molding important decisions (p. 5).

A second school of thought cites the inadequacies of the dissemination and utilization subsystems or practices of large-scale testing efforts as the major reason why such programs do not adequately serve or influence their target audiences.

And a third viewpoint challenges the suitability of large-scale testing programs, at least as currently operated, for serving the realities of educational policymaking. It is to this viewpoint that I will concentrate the remainder of my remarks.

Definitions about what we mean by the educational "policymaking process" are legion. I happen to like the definition used by Mann (1975) that educational policymaking is a process characterized by deliberations on problems which. . .

"are public in nature. . .are very consequential. . .are complex. . .are dominated by uncertainty. . .reflect and are affected by disagreement about the goals to be pursued" (p. 11).

Given this definition, let me offer three reasons why I feel our current approach to large-scale testing programs are not suited to the realities of the educational policymaking process. In so doing, I hope to stimulate some thoughts on your part about specific steps you can take to improve the impact of large-scale testing programs (and I might add, educational research and evaluation findings as well).

(1) Such programs have not adequately defined the level at which their target audiences are most likely to make policy.

Most directors of testing can tell you who they think their target audiences are; i.e., state legislators, local school board members, state or district school administrators, classroom teachers and so forth. But that is as far as their analysis has generally proceeded. They have not accounted for the fact that any one of these target audiences are likely to make policy at a variety of levels and that seldom will the same data serve these variety of needs. The literature on policy analysis or policymaking is full of definitions about different levels of policymaking. Simon (1960) makes the distinction between "organizational policy, administrative policy and operational policy" (p. 5). Others speak to distinctions between "macro-societal" policy issues where multiple institutions within a locale bear responsibility for solutions and "organizational" policy issues where a single institution is faced with the

major decisionmaking responsibility (Mann, 1975). For example, on the one hand a state board of education may be concerned with policies which relate to how the reading programs in schooling systems can contribute to equal opportunity. On the same agenda, they may be concerned with recommending whether state-approved reading textbooks should focus on "comprehension skills" or "vocabulary skills." The level of policy they are dealing with is very different and yet the statewide assessment program in reading is probably trying to assist them with both decisons using the same test scores and variables analyzed and reported in the same way. Mann made this point well when he said:

> If we are to have a way to improve. . .we must also have a way to exclude data, ignore variables, suppress interactions and focus attention on particular phenomena and relationships that are of interest. Increasing the rigor of the definition of policy, and establishing its descriptive limits, can contribute to our ability to improve the policy decision-making process (p. 7-8).

(2) Such programs seldom have the capacity to produce information which is "issue" oriented at a time when it is most needed by policymakers.

The majority of individuals responsible for large-scale testing programs with whom I have talked feel it is their primary task to produce technically sound data. They usually issue this in a report of some type soon after their analysis is completed. They feel it is somebody else's responsiblity to determine what policy issues are suggested by the data.

A few directors assume they also have this latter responsibility but only after the data has been collected and analyzed. I know of at least two statewide assessment programs that use interpretation panels after the data is available

and accompany their technical reports with specific recommendations about the type of policy issues which might be addressed. Both of these viewpoints seem to me to ignore the realities of the educational policymaking process and are a major factor as to why such data has little impact. Again, turning to the literature on policymaking, one finds general agreement that it is only at the point that a problem, demand or need becomes a publicly recognized "issue" that it is likely to receive any type of policy action (Easton, 1965; Jones, 1970; Truman, 1962). And it is at this point that decison-makers are most receptive to objective analysis of data which bears on that topic. How much more efficient and effective our large-scale testing programs would be if they could (1) identify the most consequential "issues" which their target audiences are likely to be facing and design their programs to produce information specific to those topics; (2) move away from the one-shot, once-a-year reporting syndrome and instead build a capacity to deliver the information when it is "timely" to the policymaking needs of the target audiences; and (3) develop the capacity to provide quick turn-around analyses and studies responsive to unanticipated demands which have suddenly risen to the "issue" stage. This latter capacity is not all that unrealistic to expect when you consider all of the other data collection and analysis capacities within an organization which seldom, if ever, are coordinated with the student testing program. If sampling strategies, geographic reporting boundaries and other such technical considerations were carried out in a coordinated way within most of our institutions, we could be using our student performance data far more effectively for dealing with "issue-related" decisions on a more timely and responsive schedule.

(3) <u>Few programs take into account the fact that the policymaking process is characterized by "uncertainty" and also by "competing" value systems</u>.

To put it another way, very few large-scale testing programs now have the capacity to assist their policymakers identify and analyze the consequences of alternative decisions or to do follow-up research to answer questions prompted by the original assessment results. In most of our programs, we can answer the question of "what" the student performance is in a particular area, but we certainly cannot answer the question of "why" a particular result occurred. We are, in fact, contributing to the "uncertainty" about an issue. Given scarce resources, might it be more effective to concentrate on a few critical subject areas and release existing dollars for follow-up research to provide better answers about "why" some of the data turned out the way it did? If interpretation panels are going to be used to determine the potential policy impact of assessment data, might it not be more effective to have such groups prepare a series of alternative recommendations which includes their best estimate of the educational, fiscal and political consequences of each alternative?

In summary, with the best of intentions and at great personal and fiscal cost, we have initiated large-scale testing programs for the purpose of producing useful information for target policymaking audiences. We have evidence that our efforts have missed the mark, and it is my belief that a contributing factor is our lack of understanding of and responsiveness to the practicalities and realities of the policy process itself. I have suggested at least three areas in which improvements should be made. Let me leave you with the following poem. It is supposedly anonymous, but I think it actually was written by the director of a large-scale testing program.

If you run very hard
With great effort and strain
You may clamber aboard
The last car, the wrong train.

REFERENCE LIST

Bleecher, H. The Authoritativeness of Michigan's Educational Accountability System. The Journal of Educational Research. Vol. 69, Number 4, December 1975.

deJung, J. Inventory of Standardized Tests Used During the 1972-73 School Year in Oregon Public Schools. Oregon State Department of Education, Salem, March 1973.

Easton, D. A Framework For Political Analysis. Prentice-Hall, Englewood Cliffs, 1965.

Hall, M. J. Statewide Assessment of Student Performance: A Comparative Survey. Paper to annual meeting of the American Educational Research Association, Washington, D.C., 1975.

Hawke, Sharryl, et al. State Accountability Activities and the Social Studies: A Nationwide Survey, A Proposed General Accountability Model and Some Guidelines. Social Science Education Consortium, Boulder, 1975.

House, E. R., Rivers, W., & Stufflebeam, D. An Assessment of the Michigan Accountability System. Phi Delta Kappan. Vol. 55, Number 10, June 1974.

Institute of Social Research. The Use of Social Science Knowledge in Policy Decisions at the National Level. The University of Michigan, 1975. Quoted in Education Daily, January 13, 1976.

Jones, C. O. An Introduction to the Study of Public Policy. Wadsworth Publishing Company, Belmont, 1970.

Lindbloom, C. The Policy-Making Process. Prentice-Hall, Inc., Englewood Cliffs, 1968.

Mann, D. Policy Decision-Making in Education. Teachers College Press, New York, 1975.

Simon, H. A. Administrative Behavior. Second Edition. The MacMillan Company, New York, 1960.

Truman, D. B. The Governmental Process: Political Interests and Public Opinion. Fourth Edition. Knopf, New York, 1962.